Supplementary Addendum

NLS–72 Postsecondary Education Transcript Files

Data User's Manual for Revised Transcript, Term, and Course Files

Office of Research
Office of Educational Research and Improvement
U. S. Department of Education

Supplementary Addendum

NLS–72 Postsecondary Education Transcript Files

Data User's Manual for Revised Transcript, Term, and Course Files

Clifford Adelman
Division of Higher Education
Office of Research
Office of Educational Research and Improvement

U.S. Department of Education
Lamar Alexander
Secretary

Office of Educational Research and Improvement
Diane Ravitch
Assistant Secretary

Office of Research
Milton Goldberg
Director

June 1992

Contents

Documentation

Supplementary Addendum to NLS-72 Postsecondary Education Transcript Files Data User's Manual

For Revised Transcript, Term, and Course Files

I. Background

In the summer of 1984, the National Center for Education Statistics (NCES) began the collection of postsecondary transcripts for all participants in the National Longitudinal Study of the High School Class of 1972 who had indicated on any of the NLS-72 surveys that they had attended a postsecondary school since high school graduation. The process of requesting and gathering the transcripts took more than 6 months, and is described in the original Data File User's Manual (Jones, Baker, and Borchers 1986). The transcripts were coded, placed on tape, and released to the public in August of 1986, along with the Manual.

This collection and coding was a historic undertaking: nothing like it had ever been carried out on a national scale. Even more significant in light of continuing policy debates on retention and completion in postsecondary programs is that the timeframe (1972-84) provided a record of what individuals do in postsecondary schooling through the age of 30-31. Hence, what appeared to be a college dropout at age 20 sometimes proved to be a completion at age 28 (often in a different institution in a different state).

In September of 1987, NCES issued an Addendum to the NLS-72 Postsecondary Education Transcript Files (NLS-72 PETS) Manual. This Addendum was prepared by Paula Knepper of the Longitudinal Studies Branch of NCES, and reflected her considerable work in cleaning and reducing the four (4) original data files of the NLS-72 PETS: the STUDENT file, the TRANSCRIPT file, the TERM file, and the COURSE file.

As explained in the 1987 addendum, the cleaning and reduction was necessary because the files were developed from idiosyncratic institutional records (transcripts) and not from a standardized form. Both Knepper and others who had worked with the original NLS-PETS files had found a number of "bothersome problems," principally related to credit counts, duplicate transcripts, transfer courses, term types, institutional type identification, and missing dates and grades. Among many helpful revisions, Knepper deleted unproductive and misleading variables from the STUDENT file, added dummy records and an institutional type variable to the TRANSCRIPT file, corrected some "sortdates" and term types in the TERM file, and standardized course credits and numerical course

grades in the COURSE file. She also provided instructions for variables that require special care, and an example of SAS programming that would assist users in accessing, merging, and using the NLS-PETS files.

The product of these efforts consists of four revised files, which have the SAS extensions STUDENT3, TRANS3, TERM3, and COURSE3. These are the official NCES data files.

Knepper's 1987 Addendum should be read, in concert with the original (1986) Data User's Manual, by all users of the NLS-PETS.

In a universe of over 24,000 records with 485,000 course entries and a host of other variables, any cleaning and reduction effort is bound to be incomplete, and only subsequent work with the database could reveal other problems. As Knepper wrote in the 1987 Addendum,

> "These files do not represent a totally clean database. Additional variables may later require corrections or changes, and these files will again be modified to reflect those changes and made available to researchers . . ."

Indeed, intense subsequent work with the NLS-PETS file revealed many other problems with the data. Some of these problems were residual; others arose from wholly new types of questions that were being asked of the database.

II. Nature and Rationale for a Supplementary Addendum

This addendum describes the nature, rationale, and results of the second cleaning and revision of the NLS-PETS data. The project was carried out over a period of nearly 3 years by the Office of Research (OR)/OERI, with additional support from the National Science Foundation and the U.S. Department of Labor, and NCES.

The STUDENT file was left unaltered. The other three files underwent considerable alteration. The result, four datasets with the SAS extensions STUDENT3, TRANS4, TERM4, and COURSE4, include all of Knepper's prior work, but are not official NCES datasets. Nonetheless, NCES plans to issue this version of the NLS-PETS on CD-ROM, sometime in 1992, and to use both the lessons of this revision and its new course and degree coding schemes in processing postsecondary transcripts for the High School and Beyond/Sophomore Cohort 10-year transcript sample (scheduled for completion in 1993).

The method and process of review of data quality carried out by OR were somewhat different than those employed earlier. NCES justifiable concern, reflected in the 1987 Addendum, was timeliness (i.e., making a database cleaned of gross anomalies accessible

to researchers as soon as possible). The virtue of both OR and NCES processes, though, is a high degree of reliability, since, in each case, one person was responsible for all cleaning and recoding.

The COURSE File: A Special Case

OR was more leisurely in its approach. The first problem addressed was the COURSE file and its coding scheme. The file had been coded using the 1985 version of the Classification of Instruction Programs (CIP). The basic problem with using the taxonomy of the CIP was that it was designed to classify degree or credential programs, not courses. Hence, for example, there was one and only one code for all of Economics, when we had to answer questions about the kind of Economics studied by college students. There was one and only one code for each foreign language, when we were called upon to answer questions about the percentage of college students who had indicated proficiency in a foreign language by studying at advanced levels. The CIP codes for Mathematics were absolutely unusable to describe course-taking, and, indeed, over 80 percent of the 28,000 instances of course-taking in math originally had been classified by the contractor as "Mathematics: Other."

The process of revising the taxonomy and the decision rules for recoding the COURSE file have been described at some length in A College Course Map: Taxonomy and Transcript Data (Adelman 1990). It was to this effort, in particular, that NSF and the Department of Labor contributed. Over 50 external reviewers, including teams of faculty in particular disciplines and other postsecondary curriculum experts, reviewed and contributed to the various stages of revision of the course-coding scheme that is reflected in COURSE4. For COURSE4, the taxonomy in A College Course Map supersedes the taxonomy in the official Data User's Manual. For example, there are 17 codes for courses in Economics, not one. There are 18 codes for courses in Mathematics, which are numbered so as to distinguish among precollegiate/remedial/developmental math topics, introductory college-level math courses, calculus and advanced mathematics, and vocational math. Almost all of the foreign languages have two codes each, one for introductory and intermediate courses, the other for advanced and literature courses.

We estimate that the COURSE4 file is probably 85-90 percent accurate. The problems with the remainder lie principally in contextual factors. To recode courses involved stripping titles out of a merged COURSE3/TRANS3 file, along with student major (where known), institution, and credits. We estimate that over 350,000 cases were examined in this manner, and that, of those, nearly half were recoded.

But stripping titles out of the COURSE file does not always provide enough information to tell you where something should be classified. Only the whole student record provides the context for judgment in marginal cases. For example, a course entitled "Golden Age of Russian Lit" is classifiable as "Russian: Advanced and Literature" (code 160422) only if

the student record shows other Russian language courses; otherwise, it's "Comparative Literature: European Lit in Translation" (code 230301). We deal with best guesses based on context, but one sees context only in a complete printed out record. If those records reveal mistakes, then each instance must be corrected with a line of SAS programming (e.g., "If STUID=676334 and TRANSNUM=1 and CRSENAME='GOLDEN AGE OF RUSSIAN LIT' then CRSECIP=230301"). Over 4,000 individual course entries were corrected in this manner. It was tedious business, with a diminishing rate of return.

Even after both methods were used in recoding courses, we were still left with nagging problems on the COURSE file; namely, those involving credits and grades (see below).

Identifying Problem Areas: Reading Records

Once the basic COURSE file recoding was complete, we could address other problems, such as the accuracy of the "major" designated on a transcript (if, indeed, any major had been designated). At this point, Nabeel Alsalam of NCES designed a program to configure the entire NLS-PETS for a special analysis file with more than 100 variables. We then created loops for new configurations of degrees, attendance patterns, and clusters of credit accumulations.

This analysis file made it possible to identify and correct problems in the database that were undetectable and inaccessible by any other means. In the course of OR's work, the complete records of approximately 10,000 of the 12,599 students for whom any transcripts had been received were printed out and read line by line. The records were selected by problem or seeming anomalies (e.g., students with no credits, students whose highest degree was the Associate's but who had earned more than 150 credits, Bachelor's degree students with missing degree dates or majors, students with Bachelor's degrees awarded by community colleges, students who had earned an Associate's degree but who, according to the data, had spent 0 months in postsecondary education).

III. Major Problems Identified and Addressed

As Knepper observed in the 1987 Addendum, when one works directly from institutional records, one deals with a range of idiosyncracies, even on the most basic of information. In the course of working on this special addendum to the NLS-PETS, it was determined that:

1) Some institutions had transcripts that rarely indicated that a degree was awarded or when, even when it was obvious from the transcript that the student earned a degree, e.g., for an undergraduate transcript, 140 credits, a 3.8 average, courses entitled "Senior Seminar" and "Senior Thesis" in the last term (with 'A's in both of them), and obviously fulfilled distribution requirements. Transcripts from Rutgers, Northeastern, Southern Illinois, the entire California State University (all

19 campuses), the Oregon state system (with the exception of the Eugene campus of the University of Oregon), and Ferris State (Michigan) were persistent sources of this problem. If the instances had been more scattered, we might have given the registrars the benefit of the doubt. There are over 150 Bachelor's degrees from these institutions that, in TRANS3, are "no degree." In TRANS4, they are BAs, with dates determined by the pattern of "SORTDATE" on the TERM file.

Among the guides used to resolve questions concerning undergraduate degrees conferred were:

a) The existence of transcripts from graduate or first professional programs in cases where no Bachelor's award had been indicated on any undergraduate transcript. It is possible to be admitted to some medical, dental, and veterinary schools without an undergraduate degree, usually from the junior year. But if a student with a medical school transcript, for example, had earned more than 132 undergraduate credits with an obvious major on the undergraduate transcript, it was highly unlikely that he/she did not have a Bachelor's. Three cases involving students moving from an undergraduate program to a veterinary school in the same institution were checked by phone; that is, we asked the general question, "At any time in the past 15 years, have you admitted students from your undergraduate school who had not earned a Bachelor's degree?" The answer was negative in two of those three cases. All but 10 cases under this general rubric were resolved in favor of "awarding" the Bachelor's on transcripts where it was obviously earned but not indicated.

b) The student's own responses to survey questions asked in 1979 (4th NLS Follow-Up) and 1986 (5th NLS Follow-Up) concerning the highest degree earned. The only degrees for which these questions were really valid were the Bachelor's, Master's, and Ph.D./1st Professional. Some 8,205 students out of 12,599 in the PETS sample could be checked in this manner. While survey respondents did not always tell the truth in such matters, we gave the benefit

of the doubt to any student who claimed, in both 1979 and 1986, to have earned Bachelor's or higher degrees unless the complete transcript record did not even come close to supporting the claim.

2) Some graduate degrees were also not noted on transcripts in TRANS3. If a transcript contained courses entitled "Master's Thesis" or "Ph.D. Dissertation," yet with no degree indicated, all other information on the student record was examined before a decision was made to "award" a degree. For example, if there was a "Master's Thesis" course and graduate courses taken for two or more terms subsequent to the "Master's Thesis" course, a Master's degree was entered on the TRANS4 file if none had been indicated.

   Doctoral degrees were more difficult cases to resolve. All other aspects of the transcript appearing normal, the course title, "Dissertation Defense" with a passing grade got you a degree in the following term in TRANS4 if none had been indicated on the TRANS3 version, particularly if you participated in the 5th (1986) Follow-Up survey and listed your occupation as college professor and your highest earned degree as a Ph.D. That was an easy decision. More difficult, but still resolvable, would be the case of a transcript showing six "Dissertation Research" semesters in microbiology ending in 1981, but no degree on the transcript (which was collected in 1984). If, in the 1986 survey, the individual claimed to have earned a Ph.D., he/she received the benefit of the doubt. The only open issue would be the date of the degree (see #4 below).

3) The definition of "First Professional Degree" used in the revised NLS/PETS dataset is different from that of the Integrated Postsecondary Education Data System (IPEDS) and its predecessor, the Higher Education General Information System (HEGIS). In this dataset, first professional degrees are those that usually require the prior receipt of a Bachelor's degree. They are thus post-baccalaureate degrees. In this dataset, first degrees in architecture, engineering, pharmacy, theology, accounting and chiropractic, for example, are not first professional degrees, even though they all may lead to licensure to practice a profession. Any first degree in these fields that was coded as a first professional degree in TRANS3 has been recoded as a Bachelor's degree in TRANS4. The cases were exclusively in chiropractic (7) and pharmacy (2).

   If these changes are not made, these degrees appear in the aggregate category of "Doctorate/First Professional" to be the equivalent of Ph.Ds--and earned within 4 or 5 years of high school graduation. While this distortion has been built into national data collection systems for three decades, its impact is greater on transcript analyses. Hence, the correction.

4) There were 23 cases in TRANS3 where degree dates had been coded as the date of first entry to the institution. In all these cases it was easy to enter a "best guess" date for the actual award of the degree--a minimum of one month after the last term date on the transcript. But even here, one had to look carefully at the conventions used on the transcript. If the institution signified its term dates with the first month of the term (e.g. January), and the term was a semester, then one could legitimately estimate the date of the degree as May or June. If the institution signified its term dates with the last month of the term (e.g., December), and the term was a quarter, then one could legitimately estimate the date of the degree as December or January.

   Missing dates for less-than-2-year credentials such as certificates, and for graduate or first professional degrees, presented more difficult challenges for imputation. In the case of sub-baccalaureate credentials, one is guided by data on the typical length of program type (e.g., secretarial, welding) from such sources as the National Association of Trade and Technical Schools. Since the unit of measurement in these programs is the contact hour, and the length of programs is usually expressed in a statement such as "600 contact hours," we assumed a 30 contact hour week, and added the number of weeks to the entry date to determine the credential date.

   Graduate degrees presented another type of imputation problem when the degree date was either missing or (erroneously) coded as the date of entry. If the program was in medicine, law, pharmacy, business administration, or any other field in which a dissertation is not required, the date of the degree was pegged at one month following the end-date for the last term of enrollment. For a Ph.D., the key lay in the number of terms in which "Dissertation Research" appeared for course titles, the field of study, and the guidance of average time-to-degree data published by the National Research Council in its Survey of Earned Doctorates. There were very few of these instances, and they were judged case by case.

5) Some graduate and professional school transcripts noted the prior receipt of the Bachelor's degree. In these cases, the composite student record appeared to contain two earned Bachelor's degrees. There were 36 such instances in TRANS3. In TRANS4, these "duplicate BAs" have been eliminated. In 86 other cases where a student earning a Bachelor's degree had majored in two fields, the second field had been recorded as a second Bachelor's degree in TRANS3. In TRANS4, these seeming double-BAs have been eliminated.

6) Some institutions never indicated a student's major for those who earned degrees. In the TRANS3 file, student major (MAJCIP) for BAs contains a high percentage of unknowns. In the TRANS4 file, all BAs have a major. In the TRANS3 file, one-third of the Associate's degrees did not indicate "major," even if the major was

Liberal Arts and Sciences (code 240101) or General Studies (code 240102). In the TRANS4 file, virtually all Associate's degrees have been coded with a major.

7) On TRANS3, majors or fields had been coded for 2,504 transcripts out of 11,169 on which no earned degree was noted. In TRANS4, the corresponding numbers were 3,715 out of 10,660. That means that, on 1,211 transcripts where no degree was indicated, majors or fields were determined and assigned. The decision rules involving the determination of a major or field on these "no degree" transcripts were complex. They required, first, that the transcript show 30 or more earned credits (if less than 30 but more than 12, virtually all credits had to be in the same or allied fields), with 50 percent of those credits in a given field. The higher the gross number of earned credits on a transcript the lower the percentage required for a field--but in no case was it less than 35 percent. Some interdisciplinary majors (e.g., Environmental Studies/Science) had to be inferred from combinations of courses from different fields. These were matters of judgment calls.

8) There were numerous cases in TRANS3 in which the major or field indicated on the transcript bore little or no relationship to the courses actually taken. In those cases, one suspects, the major or field indicated on the transcript is what the student said on matriculation at the institution, and, for whatever reason, was never changed--no matter what courses the student subsequently took. The same rules were used in reassigning major/field as were used in cases of initial assignment. We did not count the cases, but the net results can be inferred from frequency distributions for the variables MAJCIP1, MAJCIP2, and MAJCIP3 on both TRANS3 and TRANS4.

9) There were special problems with certificates (degree type 6) for education and other majors earning teacher certification. Teacher certification is granted by states, not colleges and universities. Hence, one does not always see certification on a transcript. Some colleges record it; most do not. The same is true of licenses (degree type 5) in fields where the state is the granting authority.

10) Some institutions did not indicate term dates, so that all courses were listed under the date for the term in which the student <u>first</u> entered the institution, if a term date was indicated at all. In the TERM3 file, the transcript is left unmodified. In the TERM4 file, the courses have been assigned term numbers and dates, using entry date and degree date as a guide. This is an imputation process, but is better than allowing them to stand (e.g., a case in which a student earns a BA in 5 years, 140 credits, yet no credits before the BA because the term dates are missing). Non-traditional transcripts (e.g. those from many proprietary schools, hospital schools of nursing) frequently evidenced this problem.

The <u>net</u> changes on the TERM file for the variable, SORTDATE, were as follows:

Unknown term date resolved to known: 122 cases.
Term dates revised: 186 cases.

Each revision applies to a term, and the term, in turn, usually covers more than one course.

11) Other institutional transcripts have term dates but only one term number. Hence, it sometimes appears that a student took dozens of courses in the same term. In the TERM4 file, these single terms are broken up to match the term dates.

12) Some institutions, particularly medical schools, do not indicate credits or credit equivalences for regular courses--let alone internships, clerkships, etc. When a course has been taken and passed, but no credits are entered, subsequent analyses are distorted (e.g. we might find a significant percentage of people who earned graduate degrees without earning graduate credits). In the COURSE4 file, credits have been assigned for as many of these cases as possible. The number of credits assigned depended on the number of courses in the term at issue and the type of school.

13) Some institutions and/or programs do not give grades in either numerical or traditional letter form. In many cases, the only performance indicators were letters such as CR, S, or P. The COURSE file has two variables for grades, CRSGRADA (letter) and CRSGRADB (number). In order to compute GPA, let alone credits, it is critical to have an entry from 0-4 under the CRSGRADB variable when a student has taken and completed a course in a regular term. For undergraduate courses in COURSE4, most cases of P were translated as 2.5 and most cases of S were translated as 2.0. For graduate courses, the translations were 3.0 and 2.5, respectively. The exceptions to these assignments involved not only transfer courses or courses taken by challenge examination (these are indicated by the TERMTYPE variable) but also courses in such areas as physical education activities, personal development, basic skills, and remedial work which, by institutional policy, may have involved non-additive credits. If we have a standard value (0-4) for CRSGRADB, the credits get added (unless the student flunked the course with a standard F, in which case CRSGRADB=0). Since there was no way to second-guess institutional policies on additive credits in these courses, the Ps, Ss, and CRs, with their corresponding 99 or 999 or 1000 entries for numerical grade, were left standing in the same manner as in the COURSE3 file.

14) Sometimes duplicate transcripts are generated within a state system or sub-system. A student attends the University of Minnesota, for example, and for some unknown reason a second transcript appears in the University of Minnesota

Central System Office. The second transcript does not look exactly like the first; but on closer inspection, it turns out to be a duplicate. When that happened, we retained the campus transcript and deleted the Central System Office version from the TRANSCRIPT, TERM, and COURSE files. In this process, it was important to guard all courses so that the campus transcript would not have any flags indicating transfer courses from the Central System version. Remember that transfer courses are only counted once, and, in this case, you want to be sure they are counted only on the retained transcript. There were three cases of duplicate transcripts.

15) There were three cases in which students were duplicated on the basis of duplicate transcripts in the same system. All three cases occurred in the Cuyahoga Community College system in Cleveland. For some unknown reason, the duplicate transcript was attached to a "student" with a different identification number. The transcripts were identical, the background characteristics of the two "students" were identical--only the student ID numbers were different. STUDENT3 reports 12,602 students in the the PETS sample. There are actually 12,599.

16) Either institutional name (FICENAME on the TRANS file) or institutional type were missing for 229 institutions from which transcripts were received. We were able to reduce the number of unknowns in this group to 7. The same problem affected over 2,000 institutions from which transcripts were requested but not received. We reduced the unknowns in this group by over half.

IV. New or Modified Variables

TRANSCRIPT File

TYPE Institutional type based on Carnegie codes for institutions in existence and coded by Carnegie in 1976. Institutions not subject to Carnegie code were first classified in a modified "Carnegie File," and then merged. Carnegie status was modified by degrees actually awarded by an institution to NLS-72 students. For example, if a branch campus of a state university was classified as a 2-year institution (code 40) in 1976 but actually awarded a Bachelor's degree to any member of the NLS-72/PETS sample, it was reclassified as a Comprehensive College. Three new Carnegie-type codes were created, and 9 existing categories were modified.

The values for TYPE are as follows:

11=Research University, Class I
12=Research University, Class II
13=Doctoral Degree-Granting Univ., Class I
14=Doctoral Degree-Granting Univ., Class II
21=Comprehensive College, Class I
22=Comprehensive College, Class II
31=Liberal Arts College, Class I
32=Liberal Arts College, Class II
40=Two-Year College (Community, Junior, Technical)
51=Theological Seminary, Other Religious Institution
52=Medical, Dental or Veterinary School
53=Other Profess. Health School (Nursing, Pharmacy, etc.)
54=Engineering/Tech (exclusively; degree-granting)
55=Business Schools (exclusively; degree-granting)
56=Schools of Art, Music, Design (exclusively)
57=Law Schools (exclusively; 4-year or graduate)
58=Teachers Colleges (exclusively; degree-granting)
59=Other Specialized Institutions (degree-granting)
60=Corporate Colleges (both degree and nondegree-granting)
65=Voc-Tech Schools, AVTIs (nondegree-granting)
68=Cosmetology Schools, Beauty Academies (nondegree-granting)
69=Other Proprietary Career Schools

ITYPE This variable aggregates the Carnegie categories and adds Central System Offices with FICE codes. The values are:

1 = Research and Doctoral Institutions. Carnegie codes 11, 12, 13, 14, plus Central System Offices of systems granting doctoral degrees.

2 = Comprehensive Colleges. Carnegie codes 21 and 22, plus Central System Offices for systems whose highest degree is the Master's.

3 = Liberal Arts Colleges. Carnegie codes 31 and 32.

4 = 2-Year/Community Colleges. Carnegie code 40.

5 = Vocational and/or Proprietary Schools awarding less than 2-year credentials. Carnegie codes 65, 68, 69.

6 = Specialized Institutions (e.g. free-standing medical schools, law schools, schools of art or music, hospital schools of nursing or radiology --not all of which are degree-granting, and corporate colleges--not all of which are specialized, but do award degrees). Carnegie codes 51-60.

8 = Specialized institutions or others that were reclassified as Comprehensive Colleges on the basis of degrees actually awarded. The bachelor's degree-granting U.S. Service academies and a number of branch campuses of state universities that had originally been classified as community colleges are included. For purposes of formatting analyses, this category should be combined with "2" above.

HBCUS Historically Black Colleges and Universities. This is a flag based on a 1979 U.S. Department of Education listing of these institutions.

BLACKINS A flag was added to indicate not only the officially designated HBCUs but also institutions in which at least 25 percent of 1976 enrollments were black.

STATE The precise state in which an institution is located is now indicated. The code range is 1-52. States are in alphabetical order, including DC, from 1-51. Code 52 covers all other jurisdictions (territories, foreign).

DEGREE Each received transcript has fields for 3 degrees (DEGREE1, DEGREE2, DEGREE3) with corresponding dates (DEGYR1. . ., DEGMON1. . .) and field (MAJCIP1. . .).

The TRANS3 codes for DEGREE1. . . were:

1 = Associate's
2 = Bachelor's
3 = Master's (including MBAs)
4 = Ph.D., 1st Professional (M.D., J.D., etc.)
5 = License
6 = Certificate
7 = No Degree
9 = Dummy (i.e., the field is empty)

TRANS4 maintains these codes, but distinguishes two kinds of "No Degree" (undergraduate and post-baccalaureate) as follows:

7 = No Degree: Undergraduate, Nondegree Post-Bacc.
8 = No Degree: 10+ Post-Baccalaureate Credits in Degree Program

This distinction enables researchers to identify those who began--but did not complete--graduate or post-baccalaureate professional degree or other programs by the fall of 1984. The credits are not necessarily in graduate or professional programs or courses, though most of them appear to be. The designation of code "8" for a degree on a transcript can be made only in the context of an entire student record (i.e., all received transcripts).

GRADFLG This is a new flag for transcripts on which the only recorded program is a graduate, first-professional, or other post-baccalaureate program, whether or not a degree was earned. There are 863 such transcripts. The flag is useful for purposes of excluding the data on these transcripts from aggregation in analyses of undergraduate education (though there are other ways of excluding such data). The flag does not cover transcripts on which both undergraduate and graduate work is recorded.

COURSE File

CNAME Carnegie Classification name of institution. Institutions not classified in the 1976 Carnegie Classification are missing this value. On the other hand, they are identified on the TRANS4 file under the variable, FICENAME.

TYPE Same as on the TRANS4 file.

CRSECIP See A College Course Map. Some 400 course codes and categories in the COURSE3 version were deleted, 338 new categories created, and 223 others modified.

V. Variables Deleted From Files

TRANSCRIPT File

GPA1, GPA2, GPA3	These fields indicated the grade point average associated with the nth degree listed on the transcript. Because of considerable corrections to the DEGREE variables and new treatments of the course grade variables (CRSGRADA and CRSGRADB) the values in these GPA fields are no longer accurate. Data users can--and should--construct their own GPA fields in analysis files.
HONORS3	Flag for honors on the third credential earned on a transcript. There were only 6 cases.
MINCIP3	CIP code for the 3rd Minor on a transcript. There were only 2 cases.
MINTEXT3	Verbatim text for the 3rd Minor on a transcript. There were only 2 cases.

VI. Sample Frequency Differences

The tables below present samples of differences in frequencies between the previous (extension "3") and current (extension "4") versions of the NLS/PETS files. There are far more courses and course codes (crsecip), more grade categories (letter and numerical), and more categories of credits than indicated in these tables. The intent is to be illustrative, not exhaustive.

COURSE FILE		COURSE3	COURSE4
CRSECIP	and Category Name		
999999	Unknown	4789	5818
908000	Unclassifiable Title	----*	1102
903000	Orientation: freshman, field	----*	1099
010103	Agricultural Economics	278	301
010506	Horseshoeing	8	----*
020402	Agronomy	94	107
030504	Forest Engineering	18	----*
030506	Forest Management/Silviculture	149	179
050201	Afro-American/Black Studies	659	623
230702	Afro-American Literature	----*	156
450817	Afro-American History	----*	290
500910	Afro-American Music	----*	48
060201	Accounting	10,042	----*
060201	Accounting: Intro., General	----*	6095
060202	Tax Accounting	----*	993
060203	Acc't: Intermed, Cost, Audit	----*	4112
070101	Accounting/Bookkeeping Support	357	154
070102	Clerical/Secretarial Accounting	93	62
070103	Bookkeeping	62	116
090101	Communications:Gen,Human,Speech	1359	4934
091001	Interpersonal/Small Grp Commun	----*	889
091002	Interviewing	----*	42
231001	Speech, Debate, Forensics	6235	----*
231001	Public Spk, Oral Interp, etc.	----*	1881
231002	Voice, Articulation, Diction	----*	270
090301	Communications Research, Theory	29	174
090401	Journalism, Reporting, Editing	1775	1111
090901	Mass Communications/Media	----*	667
090902	Public Opinion, Propaganda	----*	71
090903	Communic Law, Regulation, Ethics	----*	109
099999	Communications: Other	2884	335

		COURSE3	COURSE4
131314	Education: Physical Education	2243	1768
171001	HPER: Intro, General, Concepts	----*	365
171002	Athletic Injuries, Training	----*	157
171003	Adaptive Physical Education	----*	122
120204	Umpiring	38	----*
310501	Sports Officiating	----*	215
310601	Recreation Leadership	----*	80
310901	Recreation Practicum	----*	76
160905	Spanish (All)	4670	----*
160905	Spanish: Elementary, Intermed.	----*	4071
160925	Spanish: Advanced, Literature	----*	549
160935	Spanish for Native Speakers	----*	39
050107	Latin American Studies	173	199
050203	Hispanic-American Studies	236	290
450833	Latin American History	----*	215
142801	Textile Engineering	3	----*
150609	Textile Technology	8	----*
190901	Textiles & Clothing (Home Ec)	301	329
190903	Textiles & Clothing: Retail	25	----*
080102	Fashion & Apparel Marketing	158	133
190904	Textile Science, Fibers, Evaluat.	198	274
190999	Textiles & Clothing: Other	145	76
230401	English Composition (All)	12,021	17,010
230402	Advanced Grammar; Speak/Writ.	----*	467
230501	Creative Writing	619	784
231101	Technical & Business Writing	1123	899
070901	Business English	----*	931
232001	Remedial Writing	----*	2984
239999	English: Other	13,686	2257

* Either code and category does not exist in this version or the category covers different course titles.

Course Letter Grades (selected)

"A"	113,171	113,200
"B"	125,969	126,003
"C"	89,774	89,782
"D"	22,882	22,883
"F"	13,305	13,170
"I"	3,099	3,099
"AU" (audit)	598	598
"NO"	12,656	12,594
"P"	13,785	14,039
"S"	5,873	5,981
"WF"	647	794
"99" (unknown, missing)	25,174	24,814

	COURSE3	COURSE4
Course Numerical Grades (selected)		
"999"/"999.999" (none)	103,941	99,916
"4"	109,677	109,753
"3.5"	973	974
"3"	121,250	122,913
"2.5"	903	3,296
"2.3"	4,347	4,344
"2.0"	85,986	86,113
"1.0"	22,077	22,075
"0.7"	296	296
"0"	14,192	13,972
Course Credits Possible (selected)		
"999"/"999.999" (unknown, missing)	22,276	18,819
"0"	11,933	11,500
"0.5"	4,125	4,133
"1"	47,180	46,690
"2"	36,933	37,601
"3"	232,836	235,603
"4"	76,382	77,406
4.5"	2,308	2,308
"6"	5,238	5,278
"8"	1,798	1,808
"10"	711	704
"12"	730	735
"15"	309	302

TRANSCRIPT FILE	TRANS3	TRANS4
Total received in-scope transcripts	18,861	18,855
1st Earned Credential on Received Transcript		
Associate's	1,683	1,589
Bachelor's	4,816	5,103
Master's	424	473
Ph.D., 1st Professional	191	196
License	14	17
Certificate	563	578
No Degree: Undergraduate	11,169	10,660
No Degree: 10+ Post-Bacc Credits	---*	239
Dummy	1	0

	TRANS3	TRANS4
2nd Earned Credential on Received Transcript		
Associate's	36	28
Bachelor's	122	86
Master's	239	247
Ph.D., 1st Professional	50	58
Certificate	87	98
No Degree: Undergraduate	9	52
No Degree: 10+ Post-Bacc Credits	---*	211
Dummy	18,318	18,074
3rd Earned Credential on Received Transcript		
Associate's	6	2
Master's	14	8
Ph.D., 1st Professional	9	6
Certificate	14	9
No Degree: Undergraduate	3	0
No Degree: 10+ Post-Bacc Credits	---*	8
Dummy	18,815	18,822

* Category did not exist on TRANS3.

Year of First Credential on Received Transcript		
1972	43	21
1973	144	151
1974	817	829
1975	610	613
1976	2866	2905
1977	1160	1221
1978	612	659
1979	454	485
1980	380	404
1981	282	298
1982	157	165
1983	111	119
1984	52	78
Unknown	5	0
Field of 1st Credential Unknown/Indeterminable (Code 999999)		
Associate's	494	1
Bachelor's	240	0
Master's	34	0
Ph.D., 1st Professional	24	0
License	0	0
Certificate	87	2
No Degree: Undergraduate	8665	6945
No Degree: 10+ Post-Bacc Credits	---*	16

	TRANS3	TRANS4
Field of 2nd Credential Unknown/Indeterminable (Code 999999)		
Associate's	21	1
Bachelor's	72	0
Master's	127	1
Ph.D., 1st Professional	28	0
Certificate	58	2
No Degree: Undergraduate	7	9
No Degree: 10+ Post-Bacc Credits	---*	12

* Category did not exist for TRANS3.

Using the Modified NLS/PETS Files

The basic JCL for accessing the four basic files of the NLS/PETS should be modified. This is how it would read for a SAS program utilizing the NIH computer facilities with an IBM system:

```
//FODNLS JOB (WPD1,227,C),ADELMAN  [change to fit user]
/*ROUTE XEQ 9TRACKHI
/*ACCESS WPG4TOO
/*MESSAGE 113456,R
//PROCLIB DD DSN=ZABCRUN.PROCLIB,DISP=SHR
//STEP1  EXEC SAS
//STUD DD DSN=NLS.STUDENT4,UNIT=TAPE,DISP=(OLD,KEEP),
//  VOL=SER=113456,LABEL=(1,SL)
//TRAN DD DSN=NLS.TRANS4,UNIT=TAPE,DISP=(OLD,KEEP),
//  VOL=SER=113456,LABEL=(2,SL)
//TERM DD DSN=NLS.TERM4,UNIT=TAPE,DISP=(OLD,KEEP),
//  VOL=SER=113456,LABEL=(3,SL)
//CORS DD DSN=NLS.COURS4,UNIT=TAPE,DISP=(OLD,KEEP),
//  VOL=SER=113456,LABEL=(4,SL)
```

Other computer centers and systems will need further modifications according to the hardware environment, and the file names will need to be changed to conform to the name of the copies being accessed. Remember that the internal SAS name will remain the same.

Knepper's advice in the 1987 Addendum on using these files bears repetition:

> "The PETS files are hierarchical files, and must be merged to get complete information for each student. Further, these files generate multiple records per student. For most analyses and national estimates, these multiple records need to be summarized into a single record per student with the appropriate weight."

Two additional steps for most types of analyses are suggested. First, modify the COURSE4 file to exclude not only transfer courses recorded on more than one transcript (otherwise they will be counted twice), but also courses for which the letter grades (CRSGRADA) indicated either withdrawal (W, WP, WF) or incomplete (I). There are over 50,000 such courses out of 485,000. The residual represents all courses that were attempted and completed, whether or not credit was granted, whether or not the student received a passing grade. Call this new COURSE file something other than COURS4 (e.g. COURSES), and replace the appropriate line of the JCL with it.

Data users do not have to follow this suggestion to the letter. Researchers may want to exclude more (e.g. courses with letter grades E and F so that you wind up with passed courses only) or

less (you may wish to analyze transfer courses in the context of full records). But in either case, a modified COURSE file would need to be created before you built a full, merged analysis file.

Second--and this is a far more difficult task--we suggest creating a preliminary analysis file that will help you identify groups of students you want to know more about, and examine their full records so that you can better think through the implications of other variables you will want to create for your final analysis file. The final analysis file should be designed to merge with other data from the NLS-72 archive.

This preliminary file must create critical variables both within and across transcripts, variables such as highest degree earned, aggregate undergraduate credits, aggregate total credits, type of institution awarding a degree, date of entry to postsecondary education, and number of missing transcripts (if any) in a student's total record. Using this preliminary analysis file, you then might want to assemble and print out the complete records of groups of students who meet certain criteria in order to understand what other variables you need to create.

Assembling and printing out such records on a mainframe is a very expensive proposition. If you have PC-SAS and a local printer with a wide carriage, it will obviously cost much less. Below, for example, is the mainframe SAS program we used to print out the full records of NLS-PETS students who (a) entered postsecondary education directly within six months of high school graduation in 1972, (b) whose highest earned degree by the fall of 1984 was an Associate's, (c) who earned the Associate's degree from a community college, yet (d) also earned 150 or more undergraduate credits over the 12-year period (far in excess of what was required for an Associate's degree).

In the JCL for this program (which obviously does not refer to the tape copy), you will notice that we draw on a preliminary analysis file called STUPETS (actually this file is more than "preliminary") that incorporates both a modified COURSE file and the STUDENT3 file. STUPETS made possible the identification of highest degree, type of institution awarding degree, etc.:

```
//FODPRINT JOB (WPD1,227,C,300,100),ADELMAN [change to fit user]
/*UNNUMBERED
/*XEQ MSS
//PROCLIB DD DSN=ZABCRUN.PROCLIB,DISP=SHR
// EXEC SAS
//TRAN DD DSN=WPD1FOD.TRANS4,UNIT=MSS,DISP=SHR
//TERM DD DSN=WPD1FOD.TERM4,UNIT=MSS,DISP=SHR
//CORS DD DSN=WPD1FOD.COURSE4,UNIT=MSS,DISP=SHR
//STUP DD DSN=WPG1NUA.STUPETS,UNIT=MSS,DISP=SHR

options noovp nocenter;
```

```
DATA SHOWME(keep=stuid);
  set STUP.STUPETS(rename=(idlong=stuid) keep=idlong tcred tcredb
  deg4 deg5 hdeg majcip4 majcip5 numtrans msstrns psbeg psend
  type4 type5 f2yr f4yr l2yr l4yr);

  *** Who do we want? ***

  If 71<psbeg<73 and hdeg=4 and type4=4 and tcredb=>150;

DATA Step1;
  Merge SHOWME(in=inshowme) TRAN.TRANS4(keep=stuid transnum fice
    degree1 degree2 degree3 degyr1 degyr2 degyr3 degmon1 degmon2
    degmon3 majcip1 majcip2 majcip3 ficename findisp);
  by stuid;
  IF inshowme;

DATA Step2;
  Merge STEP1(in=instep1) TERM.TERM4(keep=stuid transnum termnum
    termtype sortdate transfer);
  by stuid transnum;
  IF instep1;

DATA Step3;
  Set CORS.COURSE4;
  Proc Sort; by stuid transnum termnum;

DATA Step4;
  Merge STEP2(in=instep2) STEP3(in=instep3);
  by stuid transnum termnum;
  IF instep2;

PROC PRINT data=STEP4;
  VAR fice transnum transfer termnum sortdate crsecred crsecip
  crsename crsgrada crsgradb degree1 degyr1 degmon1 degree2
  degyr2 degmon2 degree3 majcip1 majcip2;
  by stuid;
  Pageby stuid;
```

There are other ways to print out the data in STEP4; and one can always ask for more (or fewer) variables than the 19 indicated. It depends on what one needs.

Other information relevant to the analysis of postsecondary careers does not exist on any of the PETS files, but can be gleaned from the NLS-72 survey files. Topics such as change of major, use of Advanced Placement and credit-by-examination, scholarships and loans, aspirations and plans, and satisfaction with various aspects of postsecondary education, are all covered in the surveys. Users will need to construct a separate analysis file utilizing appropriate variables from the surveys and will need to merge this file with the transcript data in order to perform more complete analyses.

FILE CONTENTS: STUDENT4

DSNAME = NLS.STUDENT4
NUMBER OF OBSERVATIONS: 14,759 NUMBER OF VARIABLES: 18

ALPHABETIC LIST OF VARIABLES AND ATTRIBUTES

#	VARIABLE	TYPE	LENGTH	POSITION	LABEL
3	BYSES	NUM	8	11	BY SES COMPOSITE
17	BYSEST	NUM	2	69	BY SES TRICHOTOMY
16	BYTEST	NUM	2	67	BY TEST TRICHOTOMY
12	CRACE	NUM	2	59	COMPOSITE RACE
13	CSEX	NUM	2	61	COMPOSITE SEX
18	HSGRADES	NUM	8	71	HIGH SCHOOL GRADES
14	HSPGM	NUM	2	63	HIGH SCHOOL PROGRAM
10	NUMTRANS	NUM	2	55	# OF TRANS RECEIVED
1	PARTIC	NUM	3	4	PARTICIP. FLAGS
8	RECTYPE1	NUM	2	51	RECORD TYPE INDICATOR
15	REGION	NUM	2	65	H.S. REGION
9	REQTRANS	NUM	2	53	# OF TRANS REQUESTED
2	STUID	NUM	4	7	STUDENT ID #
11	TRNSFERS	NUM	2	57	TRANSFER COURSES FLAG
5	WT1	NUM	8	27	TRANSCRIPT WEIGHT
6	WT2	NUM	8	35	FU4 WEIGHT
7	WT3	NUM	8	43	PANEL WEIGHT
4	WT7	NUM	8	19	NLS BASE WEIGHT

FILE CONTENTS: TERM4

DSNAME = NLS.TERM4
NUMBER OF OBSERVATIONS: 120,885 NUMBER OF VARIABLES: 14

ALPHABETIC LIST OF VARIABLES AND ATTRIBUTES

#	VARIABLE	TYPE	LENGTH	POSITION	LABEL
12	COURSES	NUM	2	55	# OF CRSES THIS TERM
14	GRADCODE	NUM	2	59	GRADE SCALE TYPE
5	HIGHGRAD	NUM	3	16	SCALE HIGHEST GRADE
6	LOWGRAD	NUM	3	19	SCALE LOWEST GRADE
7	MEMO	CHAR	25	22	MEMO/NAME OF TRANSF SCH
4	PASSGRAD	NUM	3	13	PASSING GRADE (NUM)
8	RECTYPE3	NUM	2	47	RECORD TYPE INDICATR
13	SEASON	NUM	2	57	SEASON OF THIS TERM
3	SORTDATE	NUM	3	10	DATE OF TERM (YYMM)
1	STUID	NUM	4	4	STUDENT ID #
9	TERMNUM	NUM	2	49	TERM # ON TRNSCRPT
11	TERMTYPE	NUM	2	53	TYPE OF TERM
10	TRANSFER	NUM	2	51	TR CRSES IN TRNSCRPT
2	TRANSNUM	NUM	2	8	TRANSCRIPT NUMBER

FILE CONTENTS: COURSE4

DSNAME = NLS.COURS4
NUMBER OF OBSERVATIONS: 484,522 NUMBER OF VARIABLES: 15

ALPHABETIC LIST OF VARIABLES AND ATTRIBUTES

#	VARIABLE	TYPE	LENGTH	POSITION	LABEL
12	CNAME	CHAR	25	65	[INSTITUTION NAME]
15	CRSECIP	NUM	8	101	COURSE CIP CODE
4	CRSECRED	NUM	8	12	CRSE CREDITS POSSIBLE
6	CRSENAME	CHAR	25	28	VERBATIM CRSE TITLE
8	CRSGRADA	CHAR	2	55	LETTR GRADE FOR CRSE
5	CRSGRADB	NUM	8	20	NUMER GRADE FOR CRSE
10	FICE	NUM	4	59	SCHOOL ID
9	GRADTYPE	NUM	2	57	TYPE OF GRADE
11	INSTYPE	NUM	2	63	INSTITUTION TYPE
14	ITYPE	NUM	2	93	[ALT INSTIT. TYPE]
7	RECTYPE4	NUM	2	53	RECORD TYPE INDICATR
1	STUID	NUM	4	4	STUDENT ID #
3	TERMNUM	NUM	2	10	TERM # ON TRNSCRPT
2	TRANSNUM	NUM	2	8	TRANSCRIPT NUMBER
13	TYPE	NUM	3	90	CARNEGIE TYPE

FILE CONTENTS: TRANS4

DSNAME = NLS.TRANS4
NUMBER OF OBSERVATIONS: 24,253 NUMBER OF VARIABLES: 42

ALPHABETIC LIST OF VARIABLES AND ATTRIBUTES

#	VARIABLE	TYPE	LENGTH	POSITION	LABEL
38	BLACKINS	NUM	3	257	[PREDOM BLACK INST]
35	COFCON	NUM	8	216	INST OFFERING & CONTROL
26	DEGMON1	NUM	2	198	MO. RCVD 1ST CREDENTIAL
30	DEGMON2	NUM	2	206	MO. RCVD 2ND CREDENTIAL
33	DEGMON3	NUM	2	212	MO. RCVD 3RD CREDENTIAL
8	DEGNAME1	CHAR	7	32	VERBATIM, 1ST CRED
9	DEGNAME2	CHAR	7	39	VERBATIM, 2ND CRED
10	DEGNAME3	CHAR	7	46	VERBATIM, 3RD CRED
24	DEGREE1	NUM	2	194	1ST CREDENT EARNED
28	DEGREE2	NUM	2	202	2ND CREDENT EARNED
32	DEGREE3	NUM	2	210	3RD CREDENT EARNED
27	DEGYR1	NUM	2	200	YR. RCVD 1ST CREDENTIAL
31	DEGYR2	NUM	2	208	YR. RCVD 2ND CREDENTIAL
34	DEGYR3	NUM	2	214	YR. RCVD 3RD CREDENTIAL
2	FICE	NUM	4	8	SCHOOL ID
37	FICENAME	CHAR	25	232	[SCHOOL NAME]
20	FINDISP	NUM	2	186	FINAL DISPOSITION
42	GRADFLG	NUM	8	284	[POST-BACC PROG ONLY]

40	HBCU	NUM	8	268	[HIST. BLACK COLLEGE]
25	HONORS1	NUM	2	196	HONORS, 1ST CREDENTIAL
29	HONORS2	NUM	2	204	HONORS, 2ND CREDENTIAL
22	INSTYPE	NUM	2	190	INSTITUTION TYPE
39	ITYPE	NUM	8	260	[AGGREGATED TYPE]
3	MAJCIP1	NUM	4	12	CIP CODE, 1ST MAJOR
4	MAJCIP2	NUM	4	16	CIP CODE, 2ND MAJOR
5	MAJCIP3	NUM	4	20	CIP CODE, 3RD MAJOR
11	MAJTEXT1	CHAR	25	53	VERBATIM, 1ST MAJOR
12	MAJTEXT2	CHAR	25	78	VERBATIM, 2ND MAJOR
13	MAJTEXT3	CHAR	25	103	VERBATIM, 3RD MAJOR
6	MINCIP1	NUM	4	24	CIP CODE, 1ST MINOR
7	MINCIP2	NUM	4	28	CIP CODE, 2ND MINOR
14	MINTEXT1	CHAR	25	128	VERBATIM, 1ST MINOR
15	MINTEXT2	CHAR	25	153	VERBATIM, 2ND MINOR
21	PSREGION	NUM	2	188	POST-SEC SCHL REGION
16	RECTYPE2	NUM	2	178	RECORD TYPE INDICATOR
41	STATE	NUM	8	276	[STATE: INSTIT LOCATION]
1	STUID	NUM	4	4	STUDENT ID #
23	TERMS	NUM	2	192	# OF TERMS ON TRANSCRIPT
17	TRANSNUM	NUM	2	180	TRANSCRIPT NUMBER
19	TRNSFERT	NUM	2	184	TRANSFER CRSE FLAG
18	TRNSFLAG	NUM	2	182	RCVD TRANSCRIPT
36	TYPE	NUM	8	224	CARNEGIE TYPE

United States
Department of Education
Washington, D.C. 20208–5647

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

FOURTH CLASS BOOK RATE

OR 92–3534